Factor Analysis

MARTIN SEBERA

MAGDEBURG 15. 12. 2023

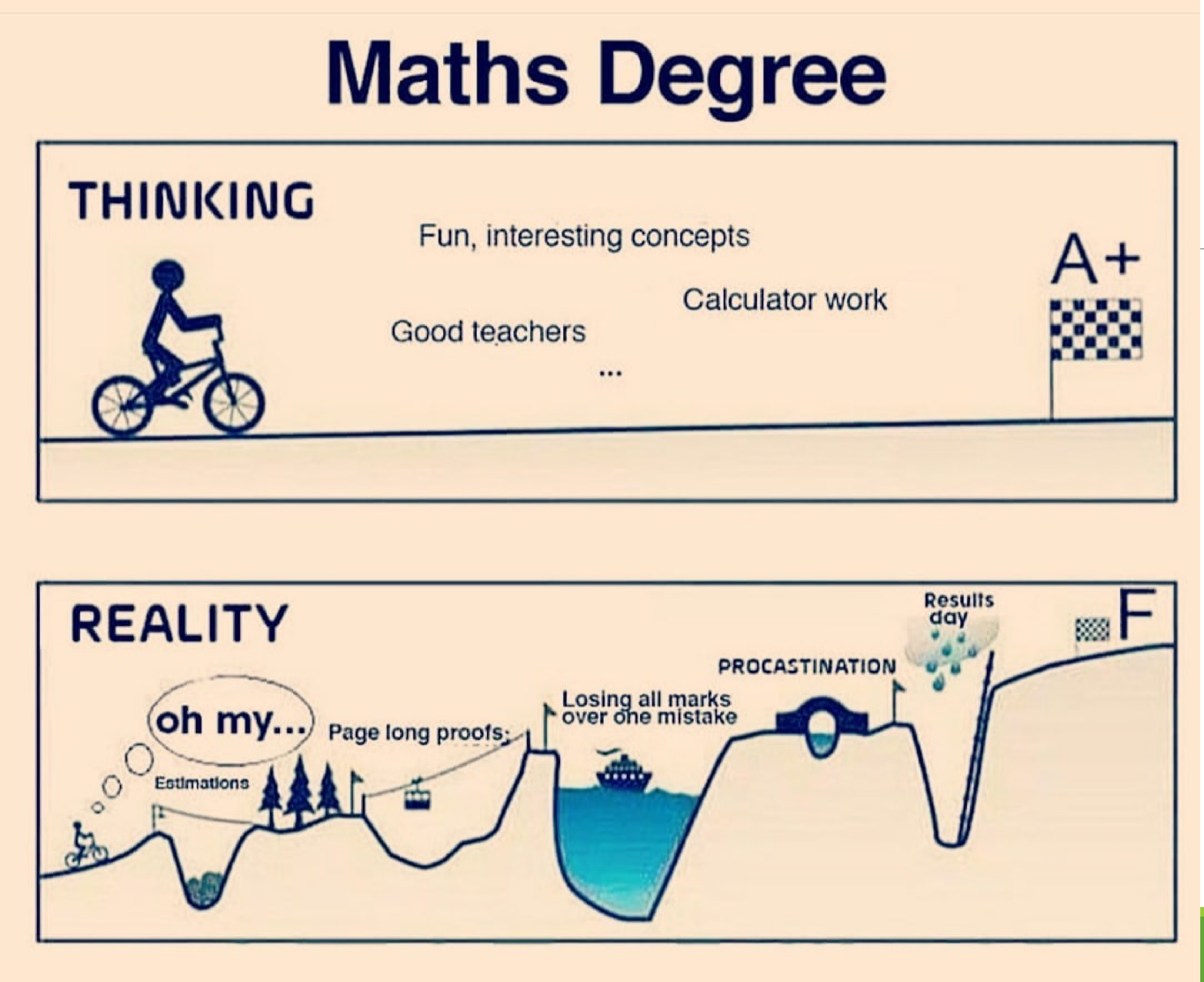
Martin Sebera

Faculty of Sports Studies, Masaryk University Brno
Czech Republic

Interests: mathematics, statistics, programming, artificial intelligence, esports

https://www.muni.cz/en/people/55084-martin-sebera

Email: sebera@fsps.muni.cz


Maths Degree
THINKING
Fun, interesting concepts
A+
Calculator work
Good teachers
...
REALITY
Results
day
F
PROCASTINATION
Losing all marks
over one mistake
oh my...
Page long proofs
Estimations

Lecture schedule – Factor Analysis

What is it used for

Requirements

Procedure

Weaknesses

Conclusion

Example - Decathlon

- sw TIBCO Statistica 14

- sw IBM SPSS 28

When was it created and who discovered it?

1. Factor analysis originated in the field of psychology.
2. Its founder is considered to be Charles Spearman
3. in 1904 in an article on the nature of intelligence proposed the hypothesis of the existence of a common factor of "general intellectual ability", causing correlations between the results of various intelligence tests

What is it used for

1. To identify groups of variables that are interrelated and can be represented by a smaller number of factors or latent variables.
2. Dimensionality reduction: Factor analysis allows data simplification by reducing many measured variables to a smaller number of factors.
3. Structure Identification: Helps identify hidden structure in a data set, which can for example reveal groups of variables that may represent a common concept or factor.
4. Data Exploration: It is useful for data exploration when researchers are looking for patterns or relationships in complex datasets.

Requirements

1. **Sample size**: at least 5-10 observations for each variable, but ideally the total sample should have at least 100 observations. Larger samples provide more robust and stable results.
2. **Linear relationships**: Factor analysis assumes linear relationships between variables.
3. **Normal distribution of data**: Although factor analysis can be performed on data that is not normally distributed, the normal distribution increases the reliability.
4. **Homogeneity of the sample**: The data should come from a homogeneous group or population so that the results of the factor analysis are relevant and interpretable for that population.
5. **No or minimal missing values**

Procedure 1/2

Data preparation:

- sample size, linear relationships between variables, normal distribution, etc.
- **Data cleaning** - including addressing missing values and removing outliers.

Method selection - Deciding whether to use exploratory factor analysis (EFA) or confirmatory factor analysis (CFA). EFA is used for discovering potential structures, while CFA for testing hypotheses about the structure.

Calculation of the correlation matrix - Creating a correlation matrix of variables. This matrix provides the basis for factor analysis.

Choosing an extraction method - principal axes (Principal Axis Factoring) or maximum likelihood

Procedure 2/2

Selection of the number of factors to be extracted, which can be done using the Kaiser criterion (eigenvalues > 1), the scree test, or based on theoretical considerations.

Factor rotation – better interpretation. Orthogonal rotation (e.g. Varimax) maintains factor independence.

Interpretation of factors - Each factor is interpreted on the basis of variables that have high loadings on it. Interpretation depends on the research context.

Assessment of model fit and reliability - In CFA, model fit is evaluated using various measures such as RMSEA (Root Mean Square Error of Approximation), CFI (Comparative Fit Index), and others.

Important terms again - 1/3

The principal component method gives uncorrelated factors, which are additionally ordered according to their variance, such that the first factor has the largest variance and the last the smallest.

Factor analysis can be considered as its extension.

While principal component analysis tries to reduce the number of variables so that the variance of the original variables is best clarified, factor analysis tries to clarify the correlations of the original variables as best as possible.

Important terms again - 2/3

Factor rotation

1. there are infinitely many factor solutions.
2. The factors are transformed so that we can interpret them as best as possible.
3. At the same time, practice has shown that factors whose factor loadings take on values close to either one or zero are best interpreted

Important terms again - 3/3

Interpretation of factors

1. We describe a factor as having something in common in content with those variables that have high factor loadings on that factor.
2. When interpreting the factors, one must be careful and think about whether the name of the factor is really behind its real existence.
3. If it does not have a logical explanation for the factor, we cannot use factor analysis

Weaknesses

1. Complexity and subjectivity: The interpretation of factors can often be subjective and depends on the researcher's decisions (e.g. choice of number of factors, rotation).
2. Assumptions about the data: linear relationships between variables and normal distribution, which do not always correspond to the actual data.
3. Dependence on sample size: A large enough sample is needed for reliable results.
4. Limitation to linear relationships: Factor analysis cannot effectively handle non-linear relationships between variables.
5. Unclear meaning of factors: Identified factors may not always have a clear or intuitive meaning and may require further research to be fully understood.

Conclusion

It is always important to remember that

1. no statistical technique is all-powerfull,
2. it is necessary to evaluate the appropriateness of the method in relation to the data and objectives of your research.
3. If you are unsure, it may also be helpful to consult an expert in statistics or research methodology about the issue.

Example - Decathlon results from the Athens Olympics 2004

1. sw TIBCO Statistica 14
2. sw IBM SPSS 28

Example - Decathlon results from the Athens Olympics 2004

1. Objective of the analysis - We are interested in whether it is possible to identify the factors on which the results in individual disciplines depend. And further which factors are most important for victory.

2. Data standardization - Not necessary. Factor analysis, as a method based on the correlation matrix, is not dependent on the scale of the input values.

3. Factor estimation methods - the method of principal components,

	1 Team	2 Finished	3 Points	4 100 m	5 Long jump	6 Shot put	7 High jump	8 400 m	9 110 m hurdles	10 Discus Throw	11 Pole vault	12 Javelin	13 1500 m
Andreev Pavel	101	0	5456	11,29		14,3	2	51,64	15,54	41,89	4,9		
Dvorák Tomáš	102	0	746	11,53									
Leyckes Dennis	103	0	3056	11,05	7,05	12,84	1,91						
Llanos Luiggy	104	0	5737	10,94	7,43	13,77	1,91	49,28	14,13	41,82			
Lobodin Lev	105	0	1631	11,05	6,86								
Magnússon Jón Arnar	106	0	2480	11,05	7,12	14,98							
Moussa Ahmad Hassan	107	0	3936	10,79	7,04	13,32	1,82	48,73					
Pappas Tom	108	0	6182	10,8	7,38	16,17	2,03	47,97	14,18	47,39			
Rahnu Kristian	109	0	1668	10,77		14,45							
Averyanov Nikolay	105	1	8021	10,55	7,34	14,44	1,94	49,72	14,39	39,88	4,8	54,51	271,02
Barras Romain	110	1	8067	11,14	6,99	14,91	1,94	49,41	14,37	44,83	4,6	64,55	267,09
Bernard Claston	111	1	8225	10,69	7,48	14,8	2,12	49,13	14,17	44,75	4,4	55,27	276,31
Casarsa Paolo	112	1	7404	11,36	6,68	14,92	1,94	53,2	15,39	48,66	4,4	58,62	296,12
Clay Bryan	108	1	8820	10,44	7,96	15,23	2,06	49,19	14,13	50,11	4,9	69,71	281,65
Covalenko Victor	113	2	6543	11,28	7,2	13,04	1,85	51,82	15,8	38,19		53,46	263,81
Drews Stefan	103	1	7926	10,87	7,38	13,07	1,88	48,51	14,01	40,11	5	51,53	274,21
Gómez David	114	1	7865	11,08	7,26	14,57	1,85	48,61	14,41	40,95	4,4	60,71	269,7
Hernu Laurent	110	1	8237	10,97	7,19	14,65	2,03	48,73	14,25	44,72	4,8	57,76	264,35
Karlivans Janis	115	1	7583	11,33	7,26	13,3	1,97	50,54	14,98	43,34	4,5	52,92	278,67
Karpov Dmitriy	116	1	8725	10,5	7,81	15,93	2,09	46,81	13,97	51,65	4,6	55,54	278,11
Korkízoglou Pródromos	117	1	7573	10,86	7,07	14,81	1,94	51,16	14,96	46,07	4,7	53,05	317
Lorenzo Santiago	118	1	7592	11,1	7,03	13,22	1,85	49,34	15,38	40,22	4,5	58,36	263,08
Macey Dean	119	1	8414	10,89	7,47	15,73	2,15	48,97	14,56	48,34	4,4	58,46	265,42
Martineau Eugene	120	2	7185	10,99	6,84		2	49,1	15,02	40	4,8	63,62	271,79
Nool Erki	109	1	8235	10,8	7,53	14,26	1,88	48,81	14,8	42,05	5,4	61,33	276,33
Ojaniemi Jaakko	121	1	8006	10,68	7,5	14,97	1,94	49,12	15,01	40,35	4,6	59,26	275,71
Parkhomenko Alexandr	122	1	7918	11,14	6,61	15,69	2,03	51,04	14,88	41,9	4,8	65,82	277,94
Pogorelov Aleksandr	105	1	8084	10,95	7,31	15,1	2,06	50,79	14,21	44,6	5	53,45	287,63

Example - Decathlon results from the Athens Olympics 2004

4. Eigennumbers & How many factors to create

There are three eigenvalues > 1 and the factors/components corresponding to them describe roughly 70% of the variability of the original variables. It is to be considered whether to use a fourth factor, which would increase the percentage of explained variance to 78%.

Eigenvalues of correlation matrix, and related statistics (Dec
Active variables only
Include condition: v2=1

Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	3,545628	35,45628	3,54563	35,4563
2	1,969494	19,69494	5,51512	55,1512
3	1,421791	14,21791	6,93691	69,3691
4	0,903646	9,03646	7,84056	78,4056
5	0,563241	5,63241	8,40380	84,0380
6	0,527759	5,27759	8,93156	89,3156
7	0,432437	4,32437	9,36400	93,6400
8	0,365718	3,65718	9,72972	97,2972
9	0,164039	1,64039	9,89375	98,9375
10	0,106246	1,06246	10,00000	100,0000

Example - Decathlon results from the Athens Olympics 2004

5. Factor loadings & Rotation and interpretation of factors

We try to achieve that each factor is correlated only with a certain group of variables and the correlations with the other variables are zero. The goal is to find meaningful factors.

We select using the Varimax method.

Factor Loadings (Varimax raw) (Decathlon)
Extraction: Principal components
(Marked loadings are >,700000)
Include condition: v2=1

Variable	Factor 1	Factor 2	Factor 3
100 m	-0,842660	-0,217283	-0,064756
Long jump	0,853894	0,167822	-0,045244
Shot put	0,185353	0,863229	0,069407
High jump	0,253925	0,741986	0,001451
400 m	-0,798454	-0,036115	0,443821
110 m hurdles	-0,708282	-0,207337	0,117338
Discus Throw	0,090322	0,850633	0,087463
Pole vault	0,475904	-0,230255	0,498277
Javelin	-0,079191	0,493663	-0,469341
1500 m	-0,224590	0,183246	0,895017
Expl.Var	2,968591	2,469252	1,499070
Prp.Totl	0,296859	0,246925	0,149907

Example - Decathlon results from the Athens Olympics 2004

6. Interpretation. The first factor is clearly related to the results of short sprints and long jump - the better the result, the higher the value of the factor. F1 Speed factor. The strongest correlations of the second factor are with all "throwing" events and the high jump. F2 „Trunk“ strength (abdominal, back, core). The last third factor is clearly correlated with longer F3 distance running, which shows that this discipline is the most different from the others.

Factor Loadings (Varimax raw) (Decathlon)
Extraction: Principal components
(Marked loadings are >,700000)
Include condition: v2=1

Variable	**Factor 1**	Factor 2	Factor 3
100 m	-0,842660	-0,217283	-0,064756
Long jump	0,853894	0,167822	-0,045244
Shot put	0,185353	0,863229	0,069407
High jump	0,253925	0,741986	0,001451
400 m	-0,798454	-0,036115	0,443821
110 m hurdles	-0,708282	-0,207337	0,117338
Discus Throw	0,090322	0,850633	0,087463
Pole vault	0,475904	-0,230255	0,498277
Javelin	-0,079191	0,493663	-0,469341
1500 m	-0,224590	0,183246	0,895017
Expl.Var	2,968591	2,469252	1,499070
Prp.Totl	0,296859	0,246925	0,149907

sw IBM SPSS 28

Basic characteristics of observed variables

Descriptive Statistics			
	Mean	Std. Deviation	Analysis N
v100_m	10,92	0,23	28
Long_jump	7,27	0,34	28
Shot_put	14,63	0,86	28
High_jump	1,98	0,09	28
v400_m	49,61	1,27	28
v110_m_hurdles	14,55	0,44	28
Discus_Throw	44,38	3,30	28
Pole_vault	4,73	0,29	28
Javelin	58,95	4,98	28
v1500_m	277,54	11,32	28

Conditions of use of factor analysis

correlation coefficients are high
we see positive and negative correlations,
there is a conditional formatting tool for that in Excel

Correlation Matrix

	v100_m	Long_jump	Shot_put	High_jump	v400_m	v110_m_hurdles	Discus_Throw	Pole_vault	Javelin	v1500_m
v100_m	1,00	-0,70	-0,37	-0,31	0,63	0,54	-0,23	-0,26	-0,01	0,06
Long_jump	-0,70	1,00	0,20	0,35	-0,67	-0,54	0,25	0,29	0,09	-0,15
Shot_put	-0,37	0,20	1,00	0,61	-0,20	-0,25	0,67	0,02	0,38	0,13
High_jump	-0,31	0,35	0,61	1,00	-0,17	-0,33	0,52	-0,04	0,20	0,00
v400_m	0,63	-0,67	-0,20	-0,17	1,00	0,52	-0,14	-0,12	-0,05	0,55
v110_m_hurdles	0,54	-0,54	-0,25	-0,33	0,52	1,00	-0,22	-0,15	-0,08	0,18
Discus_Throw	-0,23	0,25	0,67	0,52	-0,14	-0,22	1,00	-0,18	0,25	0,22
Pole_vault	-0,26	0,29	0,02	-0,04	-0,12	-0,15	-0,18	1,00	-0,07	0,18
Javelin	-0,01	0,09	0,38	0,20	-0,05	-0,08	0,25	-0,07	1,00	-0,25
v1500_m	0,06	-0,15	0,13	0,00	0,55	0,18	0,22	0,18	-0,25	1,00

Conditions of use of factor analysis

The Kaiser-Meyer-Olkin measure takes on a value of 0.58, is high and indicates the appropriateness of using factor analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,590
Bartlett's Test of Sphericity	Approx. Chi-Square	112,179
	df	45
	Sig.	<,001

Conditions of use of factor analysis

Bartlett's test of sphericity:
Test criterion value = 112.179
Number of degrees of freedom = 45
Significance (= observed significance level) < 0.001

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,590
Bartlett's Test of Sphericity	Approx. Chi-Square	112,179
	df	45
	Sig.	<,001

Conditions of use of factor analysis

Bartlett's test of sphericity:
Test criterion value = 112.179
Number of degrees of freedom = 45
Significance (= observed significance level) < 0.001

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,590
Bartlett's Test of Sphericity	Approx. Chi-Square	112,179
	df	45
	Sig.	<,001

Significance level is < 0,001, we reject H0: The correlation matrix is unity (correlation coefficients off the diagonal are zero).
Thus, the basic assumption for the use of factor analysis is fulfilled.

Conditions of use of factor analysis

All KMO values for individual observed variables are satisfactory – greater than 0.5

Inverse of Correlation Matrix

	v100_m	Long_jump	Shot_put	High_jump	v400_m	v110_m_ hurdles	Discus_ Throw	Pole_vault	Javelin	v1500_m
v100_m	2,908	1,189	0,925	-0,081	-1,095	-0,414	-0,525	-0,019	-0,218	0,623
Long_jump	1,189	3,456	1,499	-0,989	1,851	0,193	-0,521	-0,512	-0,663	-0,771
Shot_put	0,925	1,499	3,348	-1,325	1,072	-0,130	-1,225	-0,374	-0,999	-0,757
High_jump	-0,081	-0,989	-1,325	2,135	-1,077	0,296	-0,319	0,141	0,403	0,725
v400_m	-1,095	1,851	1,072	-1,077	4,744	-0,394	0,605	0,143	-0,994	-2,756
v110_m_hurdles	-0,414	0,193	-0,130	0,296	-0,394	1,641	0,096	0,081	0,049	-0,029
Discus_Throw	-0,525	-0,521	-1,225	-0,319	0,605	0,096	2,499	0,732	-0,217	-0,972
Pole_vault	-0,019	-0,512	-0,374	0,141	0,143	0,081	0,732	1,414	-0,055	-0,547
Javelin	-0,218	-0,663	-0,999	0,403	-0,994	0,049	-0,217	-0,055	1,630	1,056
v1500_m	0,623	-0,771	-0,757	0,725	-2,756	-0,029	-0,972	-0,547	1,056	3,052

Factor extraction
Principal component method

Attention!
Unlike the tables of correlation coefficients and communalities, the rows of this table are not devoted to manifest variables, but to factors

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

The first section, entitled "Initial Eigenvalues" by SPSS, contains the results of the principal components method

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

The eigenvalues are listed in the second column, which follows the column labeled factors (components).

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

The eigenvalues of the correlation matrix indicate the variance exhausted by the factor. This variance, expressed as a percentage, is shown in the third column of the tables

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

For a better idea of how much variance is already exhausted by the given number of factors, the fourth column with the cumulative percentage values of the exhausted variance is used.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

The second part of the table "Extraction Sums of Squared Loadings" gives the amount of variance extracted after factor extraction

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

The second part of the table "Extraction Sums of Squared Loadings" gives the amount of variance extracted after factor extraction

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

absolute

Factor extraction
Principal component method

The second part of the table "Extraction Sums of Squared Loadings" gives the amount of variance extracted after factor extraction

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

percentage

Factor extraction
Principal component method

The second part of the table "Extraction Sums of Squared Loadings" gives the amount of variance extracted after factor extraction

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

in cumulative percentage form

Factor extraction
Principal component method

We see that it is limited to a given number of factors, i.e. 3.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

For principal component factor extraction, it is of course identical to the first part of the table.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

In the third part of the table "Extraction Sums of Squared Loadings" the values of the exhausted variance after rotation are presented analogously.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

In the third part of the table "Extraction Sums of Squared Loadings" the values of the exhausted variance after rotation are presented analogously. We see that the first factor accounts for 30.4% of the variance

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

In the third part of the table "Extraction Sums of Squared Loadings" the values of the exhausted variance after rotation are presented analogously. We see that the first factor consumes 30.4% of the variance, the second 24.58% and the third 14.38%.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Factor extraction
Principal component method

The method of principal components gives the factors that exhaust the highest percentage of variance of all the methods used. However, the main task of factor analysis is to clarify the original correlation matrix with the help of factors, not the variance.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,546	35,456	35,456	3,546	35,456	35,456	3,040	30,401	30,401
2	1,969	19,695	55,151	1,969	19,695	55,151	2,458	24,584	54,985
3	1,422	14,218	69,369	1,422	14,218	69,369	1,438	14,384	69,369
4	,904	9,036	78,406						
5	,563	5,632	84,038						
6	,528	5,278	89,316						
7	,432	4,324	93,640						
8	,366	3,657	97,297						
9	,164	1,640	98,938						
10	,106	1,062	100,000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix

Principal Component Analysis

Finally the result!
We can now try to interpret the factors

Rotated Component Matrix[a]

	Component		
	1	2	3
v100_m	0,819	-0,243	-0,178
Long_jump	-0,849	0,179	0,080
Shot_put	-0,173	0,868	-0,023
High_jump	-0,252	0,740	-0,060
v400_m	0,864	0,009	0,299
v110_m_hurdles	0,718	-0,205	0,021
Discus_Throw	-0,076	0,856	-0,020
Pole_vault	-0,381	-0,147	0,601
Javelin	-0,005	0,421	-0,541
v1500_m	0,377	0,302	0,807

Extraction Method: Principal Component

a. Rotation converged in 5 iterations.

Rotated Component Matrix[a]

	Component		
	1	2	3
v100_m	0,819		
Long_jump	-0,849		
Shot_put		0,868	
High_jump		0,740	
v400_m	0,864		
v110_m_hurdles	0,718		
Discus_Throw		0,856	
Pole_vault			0,601
Javelin		0,421	-0,541
v1500_m			0,807

Extraction Method: Principal Component

a. Rotation converged in 5 iterations.

Loads less than ±0.4 are deleted

Compare 4 rotation methods

	Component			Component			Component			Component		
	1	2	3	1	2	3	1	2	3	1	2	3
v100_m	0,819			0,815			0,827			0,830		
Long_jump	-0,849			-0,846			-0,854			-0,858		
Shot_put		0,868			0,869			0,865			0,877	
High_jump		0,740			0,741			0,738			0,754	
v400_m	0,864			0,868			0,857			0,870		
v110_m_hurdles	0,718			0,717			0,721			0,734		
Discus_Throw		0,856			0,856			0,855			0,859	
Pole_vault			0,601			0,608			0,587			0,559
Javelin		0,421	-0,541		0,411	-0,549		0,440	-0,526		0,414	-0,519
v1500_m			0,807			0,797			0,825			0,850
ROTATION	Varimax			Equamax			Quartimax			Promax		

Transfer to sports train

1. What to train (it is not possible to train all 10 disciplines at the same time) to be the best decathlete?

Rotated Component Matrix[a]

	Component		
	1	2	3
v100_m	0,819		
Long_jump	-0,849		
Shot_put		0,868	
High_jump		0,740	
v400_m	0,864		
v110_m_hurdles	0,718		
Discus_Throw		0,856	
Pole_vault			0,601
Javelin		0,421	-0,541
v1500_m			0,807

Extraction Method: Principal Component

a. Rotation converged in 5 iterations.

Transfer to sports train

1. What to train (it is not possible to train all 10 disciplines at the same time) to be the best decathlete?
2. You can't always generalize, but it's **a 400m run** and a **shot put spin**!
3. Why? It said Mr. Váňa, the coach of the Czech decathletes Roman Šebrle (gold and silver from the Olympics) and Tomáš Dvořák (3x world champion, bronze from the Olympics). Váňa bet on the speed of execution of individual disciplines.

Rotated Component Matrix[a]

	Component		
	1	2	3
v100_m	0,819		
Long_jump	-0,849		
Shot_put		0,868	
High_jump		0,740	
v400_m	0,864		
v110_m_hurdles	0,718		
Discus_Throw		0,856	
Pole_vault			0,601
Javelin		0,421	-0,541
v1500_m			0,807

Extraction Method: Principal Component

a. Rotation converged in 5 iterations.

And **what to do** if we are looking for relationships and the conditions for classic tests known from statistics, such as linear regression, factor analysis, etc., are not met?

Data Mining Statistical Methods

C&RT trees and Neural networks

Martin Sebera

de Oliveira Abrahão, A. A., Marcos de Andrade Júnior, É., de França Ferraz, A., Kuang Hongyu, & Fett, C. A. (2022). Factor Analysis for detection of sports talent in football players. *Saúde e Pesquisa*, *15*(1), 1–12. https://doi.org/10.17765/2176-9206.2022v15n1.e9766

Jinrui Zhang, Zhiwen Zhang, Shuo Peng, Veloo, A., Bailey, R. P., & Wee Hoe Tan. (2023). Psychometric properties of the Chinese version of Sport Anxiety Scale-2. *Frontiers in Psychology*, 1–10. https://doi.org/10.3389/fpsyg.2023.1260253

Lan Zhou, Sang-Ho Lee, & Youshen Cao. (2022). An empirical analysis of sport for mental health from the perspective of a factor analysis approach. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.960255

Li, X., Chen, J., Zhan, J., & Liu, L. (2016). A Study on Sports Tourism Competitiveness Based on Factor Analysis Method. *2016 12th International Conference on Computational Intelligence and Security (CIS), Computational Intelligence and Security (CIS), 2016 12th International Conference on, CIS*, 673–676. https://doi.org/10.1109/CIS.2016.0162

Putra, M. F. P. (2022). Construct validity test of spirituality in sports test (SIST) using confirmatory factor analysis (CFA) method. *Ovidius University Annals, Series Physical Education and Sport/Science, Movement and Health*, *22*(2), 139.

Děkuji za pozornost

Thank you for your attention

Danke für Ihre Aufmerksamkeit